

Dietmar Höttecke (Hg.)

# Chemie- und Physikdidaktik für die Lehramtsausbildung

Gesellschaft für Didaktik der Chemie und Physik  
Jahrestagung in Schwäbisch Gmünd 2008

## Schweizer Bildungsstandards: Resultate der Papier- und Bleistift-Tests

Das Konsortium HarmoS Naturwissenschaften hatte von den politischen Behörden den Auftrag, einerseits ein validiertes Kompetenzmodell vorzulegen, andererseits Basisstandards vorzuschlagen. Beides konnte bzw. kann nur auf der Basis von Tests geschehen. Zu den Tests gehörte unter anderem ein Papier- und Bleistift-Test (PB-Test), der in Klassen des 6. und 9. Schuljahrs durchgeführt wurde. Im Folgenden werden die Testplanung, -durchführung und -auswertung sowie ausgewählte Testresultate beschrieben.

### Testplanung

Nachdem das Kompetenzmodell in einer ersten Version erarbeitet war (Herbst 2006), entschloss sich das Konsortium im geplanten PB-Test drei der Handlungsaspekte zu validieren: "Informationen erschließen", "Ordnen, strukturieren, modellieren" sowie "Einschätzen und beurteilen". Zwei weitere Handlungsaspekte, "Fragen und untersuchen" sowie z.T. "Entwickeln und umsetzen" wurden im Experimentierertest validiert (siehe die folgenden Beiträge von Metzger und Gut).

Der PB-Test für das 6. und 9. Schuljahr fand im April und Mai 2007 statt, ein Jahr früher als ursprünglich gedacht. Dies führte dazu, dass relativ kurzfristig zwischen September 2006 und Februar 2007 sämtliche Aufgaben entwickelt, pilotiert und übersetzt (Deutsch und Französisch) werden mussten. Da alle Klassen nicht nur in Naturwissenschaften, sondern auch in Mathematik, Unterrichts- und Fremdsprache an insgesamt zwei Halbtagen getestet wurden, gab es für den PB-Test klare Rahmenbedingungen: 10' Testzeit pro Aufgabe, Testhefte mit je zwei Aufgaben à 10' aus unterschiedlichen Themenbereichen, 4 bis 6 Aufgaben je Schüler bzw. Schülerin.

Da nur wenig Zeit für die Aufgabenkonstruktion zur Verfügung stand, konnten die Aufgaben zwar wohl in je zwei Schulklassen pilotiert sowie von Experten und Expertinnen begutachtet werden, hingegen musste auf einen eigentlichen Feldtest verzichtet werden. Die PB-Aufgaben weisen ein ähnliches Format auf wie diejenigen in PISA. Folgende Charakteristika gelten für so gut wie alle Testaufgaben:

- Ausgehend von einer Situation können sich die Lernenden während des Beantwortens der verschiedenen Teilfragen (je Situation 3 bis 8 Teilfragen, d.h. 3 bis 8 Items) einen neuen Sinnzusammenhang erschließen.
- Je Situation werden – verteilt auf die Items – verschiedene Handlungsaspekte überprüft.
- Bei den Inhalten wird eine Balance zwischen lebensweltlichem Bezug und fachlichen Grundideen angestrebt.
- Es sind kaum Routineaufgaben, bei denen die Lernenden auf Standardlösungen zurückgreifen können.

Für den PB-Test wurden für die 6. und 9. Klasse je 45 Situationen mit je 3-8 Items entwickelt, gleichzeitig wurde je Item ein detailliertes Korrekturschema erstellt. Alle Items und Korrekturschemen existieren in Deutsch und Französisch; auf den Einbezug der italienischsprachigen Schweiz musste leider aus Zeit- und finanziellen Gründen verzichtet werden. Abbildung 1 zeigt ein typisches Item, welches in der 6. und 9. Klasse eingesetzt wurde.

### Testdurchführung

Die Durchführung der Tests erfolgte in der 6. und 9. Klasse nach jeweils ähnlichem Schema, hier die Charakteristika für das 9. Schuljahr:

- Testzeit 2 x 80 Minuten im Rahmen des nationalen Tests, davon für Naturwissenschaften zwischen 40 und 60 Minuten Testzeit; zusätzlich allgemeiner Fragebogen;

Berlin: LIT Verlag, 2009

- Durchführung der Tests in Klassen unter Aufsicht der Naturwissenschaftslehrkraft, es wurden jeweils unterschiedliche Testhefte parallel, d.h. gleichzeitig eingesetzt (Testdesign mit Rotationsplan);
  - Je Schülerin bzw. Schüler insgesamt 4-6 Situationen, d.h. 2-3 Testhefte à 2 Situationen à 10 Minuten; je Situation 3-8 Items;
  - Jedes Testheft wurde von mindestens 300 Schülerinnen und Schülern in beiden Sprachregionen (deutsch und französisch) bearbeitet;
  - Insgesamt 266 Items zu den drei getesteten Handlungsaspekten; je Handlungsaspekt zwischen 54 und 126 Items.
- Zur Stichprobengröße: Im 6. Schuljahr bearbeiteten 255 Klassen mit insgesamt 4124 Jugendlichen den Test, im 9. Schuljahr 273 Klassen mit 3888 Jugendlichen.

**Aufgabentext:**

Wofür wir viel Energie brauchen: In der Darstellung ist angegeben, wofür wir vor allem Energie brauchen. Das wird gezeigt: In der Schweiz brauchen wir von 100 Teilen (T.) Energie so viele (Anzahl T.) für [...]

Erstelle eine Rangliste (Ränge 1 bis 6; 1. Rang: am meisten Energie für ...)

- ☛ Licht
- ☛ Autos, Mofas
- ☛ Waschen
- ☛ Kochen
- ☛ Heizen
- ☛ Duschen, Baden

**Korrekturschema:** Code 2: Alle Angaben in der Rangliste richtig (d.h. 1 für Autos, 2 für Heizen etc.); Code 1: Ränge 1 und 2 richtig, bei den weiteren Rängen eine Rangfolge falsch; Code 0: mehrere Fehler.

**Itemcharakteristika:** Handlungsaspekt "Informationen erschließen"; Themenbereich "Natur, Gesellschaft, Technik - Perspektiven"; von Fachleuten eingeschätztes Niveau: I für Klasse 9, Niveau II für Klasse 6.

Abb. 1: Ein Item aus dem PB-Test für das 6. und 9. Schuljahr

### Testauswertung

Die Tests wurden nach detaillierten Korrekturschemata von deutsch- bzw. französischsprachigen Hilfskräften unter Anleitung von zwei Konsortiumsmitgliedern korrigiert. Nach der Datenbereinigung mussten circa 25% der Items gestrichen werden, sei es zum Beispiel wegen Übersetzungsfehlern, Problemen bei der Testdurchführung oder Ceiling-Effekten. Während dieser und späterer Auswertungsphasen arbeiteten Fachleute aus den Naturwissenschaftsdidaktiken und der Psychometrie eng zusammen.

Anschließend erfolgte eine Rasch-Skalierung, und zwar je separat für die 6. und 9. Klasse: Analog wie bei PISA wurden Mittelwert und Standardabweichung jeweils auf 500 bzw. 100 Punkte normiert. Die Skalierung erlaubt zum einen den Schwierigkeitsgrad einer Aufgabe, zum anderen die Testleistung einer Schülerin bzw. eines Schülers auf der gleichen Skala zu quantifizieren.

Zur Veranschaulichung: Das Item in Abbildung 1 lieferte folgende Resultate: Im 9. Schuljahr belief sich die Stichprobe auf N=325; von diesen erhielten 181 (56%) den Code 2, d.h. Item vollständig korrekt beantwortet, 14 (4%) den Code 1, d.h. die Frage nur teilweise richtig beantwortet, 87 (27%) Code 0; 43 (13%) hatten das Item nicht bearbeitet. Der Schwierig-

keitsgrad des Items betrug gemäß Raschskalierung 495. Für das 6. Schuljahr lauten die entsprechenden Zahlen: N=347; 155 (45%) für Code 2, 13 (4%) Code 1, 118 (34%) Code 0, 61 (17%) keine Bearbeitung des Items; Schwierigkeitsgrad 549.

### Resultate

Insgesamt lieferten die PB-Tests im 6. und 9. Schuljahr brauchbare Resultate, das heisst: breites Spektrum von Itemschwierigkeiten sowie - trotz drop-out von Items - immer noch genügend Fragen je Handlungsaspekt, Themenbereich und Sprachregion, um statistisch gesicherte Aussagen machen zu können. Es zeigte sich allerdings, dass die Schülerinnen und Schüler die Aufgaben generell schlechter beantwortet haben, als es die Expertinnen und Experten, d.h. Naturwissenschaftslehrkräfte und Fachdidaktikdozierende, zuvor erwartet hatten.

Die Korrelationen zwischen den drei getesteten Handlungsaspekten sind in beiden Schuljahren hoch, liegen sie doch im Durchschnitt bei 0,85, mit 0,79 als tiefstem und 0,94 als höchstem Wert. Deutlich tiefere Werte weisen die Korrelationen zwischen den Themenbereichen auf: die meisten dieser Korrelationen liegen zwischen 0,65 und 0,72, die höchste Korrelation beläuft sich auf 0,74, die tiefste auf 0,59. Bei derart hohen Korrelationen zwischen den Handlungsaspekten kann man sich fragen, inwieweit mit den drei Aspekten überhaupt unterschiedliche Kompetenzen gemessen werden. Die hohen Werte kommen allerdings nicht überraschend, da auch in anderen Fächern, z.B. in Mathematik oder in den Sprachen, die jeweiligen Kompetenzen meist hochgradig korreliert sind.

Zwischen den beiden Sprachregionen, deutsch- und französischsprachige Schweiz gibt es signifikante Unterschiede: im Allgemeinen weisen die Jugendlichen der Suisse Romande bei den einzelnen Handlungsaspekten und Themenbereichen einen um 20 bis 50 Punkte tieferen Mittelwert auf. Dies kann damit zusammenhängen, dass die Aufgaben mehrheitlich von Fachdidaktikdozierenden und Lehrkräften der Deutschschweiz entwickelt worden sind.

Weitere detaillierte Analysen, z.B. bezüglich Genderdifferenzen, Korrelationen mit den mathematischen und sprachlichen Kompetenzen, Korrelationen mit dem allgemeinen Interesse an Naturwissenschaften, folgen in den nächsten Monaten.

### Basisstandards

Das Konsortium hat den Auftrag Basisstandards inkl. Referenzbeispiele vorzuschlagen. Dabei handelt es sich schlussendlich um normative Entscheide - hier allerdings abgestützt auf empirische Testresultate: So dient für den Handlungsaspekt "Informationen erschließen" das Item aus Abbildung 1 als Referenzbeispiel für den Basisstandard 9. Schuljahr, auch wenn im jetzigen Zeitpunkt nur 60% der Jugendlichen mindestens Code 1 erzielt, d.h. die Frage mehrheitlich korrekt beantwortet haben. Mit dem Setzen von Basisstandards und deren Veranschaulichung durch Referenzbeispiele besteht die Hoffnung, dass langfristig, d.h. wenn Basisstandards und neuer Lehrplan einmal eingeführt worden sind, mehr Jugendliche als jetzt dieses Item korrekt beantworten.

Für weitere Aufgabenbeispiele, Testresultate und Interpretationen sei auf den wissenschaftlichen Schlussbericht (Adamina et al., 2008) sowie auf die Websites des Konsortiums Naturwissenschaften verwiesen: <http://harmos.phbern.ch>.

### Literatur

Adamina, M.; Labudde, P.; Gingins, F.; et al. (2008). *HarmoS Naturwissenschaften+ : Wissenschaftlicher Schlussbericht*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren